DOCUMENT RESUME

ED 275 758 TM 860 675

AUTHOR Kulik, Chen-Lin C.; And Others
TITLE Effects of Testing for Mastery on Student Learning.
PUB DATE Apr 86
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Academic Aptitude; Achievement Tests; Behavioral Objectives; Effect Size; Higher Education; High Schools; *Instructional Effectiveness; Mastery Learning; *Mastery Tests; *Meta Analysis; Models; Multiple Regression Analysis; *Outcomes of Education; Time Factors (Learning)

ABSTRACT

The purpose of this study was to determine the extent to which mastery testing contributes to instructional effectiveness and to determine the size of its effects in different outcome areas, such as aptitude and instructional time. A meta-analytic methodology was used to integrate findings from experimental studies. Twenty studies examined college-level effects, and investigated pre-college effects. Each study compared students taught with a mastery requirement on tests to students taught in the same way but without a requirement of test mastery. The results showed that testing for mastery had effects on student achievement at both college and pre-college levels. The effects of mastery testing were strong on low aptitude students. A mastery requirement had the effect of diminishing individual differences in student achievement in classes and of reducing the correlation between aptitude and achievement. Results also showed that mastery testing increased the amount of instructional time. (JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED275758

Effects of Testing for Mastery on Student Learning

Chen-Lin C. Kulik, James A. Kulik,

& Robert L. Bangert-Drowns

The University of Michigan

A paper presentation at the annual meeting
of the American Educational Research Association,
San Francisco, April 1986

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Chen-Lin C. Kulik

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 860 675

Effects of Testing for Mastery on Student Learning

Testing students for mastery is not a new idea in education. The idea of requiring students to demonstrate mastery on current material before moving on to new material was a key concept in several individualized systems of instruction used in this country during the 1920s and 1930s. But it was not until the 1960s that mastery testing became an especially influential idea in education.

Bloom's (1968) article on the mastery model is now generally recognized as the classic theoretical formulation on the topic. According to Bloom, nearly all students can master nearly all the objectives in a school subject if mastery on unit quizzes is required of all, and each is given the amount of time and the kind of instruction needed on each unit. In addition, Bloom (1976) has stated that students do not have to put in much more time on school tasks to achieve mastery-level proficiency. According to Bloom, less able students need more time to reach proficiency only in the initial stages of a course. Their need for extra time vanishes after they have mastered the fundamental material.

Mastery theorists and mastery critics have disagreed about the amount of experimental evidence available to support such claims. Most reviews (e.g., Block & Burns, 1976) support the basic claim that mastery testing often raises the overall level of student performance. Critics, however, have argued that the gains are not so large as mastry theory requires. Critics have also pointed out that reviews have not assembled much data on the time requirements of mastery testing or on its effects on individual differences in performance (Arlin & Webster, 1983). Some critics believe that individual differences in learning aptitude are more enduring than mastery theorists suppose.

The purpose of this study was to determine the extent to which mastery testing contributes to instructional effectiveness. Our goal was not only to find out whether mastery testing influences instructional effectiveness but also to determine how large its effects are in different outcome areas. The study used a meta-analytic methodology to integrate findings from numerous experiments on this topic. Each of the experiments covered in our analysis compared students taught with a mastery-requirement on tests to students taught in the same way but without a requirement of test mastery.

Method

The meta-analytic approach used in this review was similar to that described by Glass, McGaw, and Smith (1981).

Their approach requires a reviewer (a) to locate studies of an issue through objective and replicable searches; (b) to code the studies for salient features; (c) to describe study outcomes on a common scale; and (d) to use statistical methods to find relationships between study features and study outcomes.

Our search procedures yielded a total of 34 studies that could be considered methodologically sound. Twenty of these studies examined college-level effects, and 14 investigated precollege effects. Sixteen of the studies were conducted in conventional lecture-discussion classrooms, and 18 were conducted in individualized classes. The studies came from journal articles, dissertations, and ERIC documents.

Nineteen variables were used to describe treatments, methodologies, settings, and publication histories of the studies (Table 1). These 19 variables were chosen on the basis of (a) an examination of variables used to describe study features in previous reviews, and (b) a preliminary examination of dimensions of variation in the studies located for this analysis.

The instructional outcome measured in all 34 studies was student learning, most often indicated on achievement examinations given at the end of a program. To quantify outcomes in achievement, we used Glass's effect-size statistic, or ES, defined as the difference between the means of two groups divided by the standard deviation of the control group (Glass, McGaw, & Smith, 1981).

Other outcome variables measured in the studies were: (a) relationship between aptitude and achievement as indicated by the correlation of these two variables; and (b) amount of time needed for instruction. For indexing the effect of mastery testing on the relationship between aptitude and achievement, we used the difference between aptitude-achievement correlations in the experimental and control groups. As an index of the effect of mastery testing on instructional time, we used the ratio of time required by the experimental group to the time required by the control group.

Results

Examination Performance

All but 2 of the 34 studies reported positive effects of mastery testing on student learning as measured by a criterion test (Table 2). And 22 of the 32 studies with positive effects reported that the difference in amount learned by experimental and control groups was great enough to be considered statistically significant. Neither of the

studies with negative results reported statistically significant differences. Overall, these lopsided box-score results strongly favor the hypothesis that mastery testing has a positive effect on student learning.

The average effect of testing for mastery was to raise student achievement scores by 0.52 standard deviations. This is clearly a substantial effect. To interpret the average effect of mastery testing more fully, it is useful to refer to areas of the standard normal curve. Approximately 70% of the area of this curve falls below a *z*-score of 0.52. We can conclude, therefore, that students from mastery testing classes performed at the 70th percentile on their examinations, whereas the students who were taught without a mastery requirement performed at the 50th percentile on the same examinations. Or put in another way, 70% of the students who were tested for mastery outperformed the average student who was not tested for mastery.

Examination Performance and Study Features

Although the effect of testing for mastery was impressive in the typical study, the size of effect varied from study to study, and we therefore carried out further analyses to determine whether different types of studies were producing different results (Table 3). Because study features were intercorrelated, these analyses made use of multiple regression methods that took into account study-feature intercorrelations as well as relationships between study features and effects. Four predictors emerged from the regression analyses as significantly related to effect size (Table 4).

The first factor--mastery level--was positively related to size of effect. The higher the mastery criterion used with the experimental group, the stronger the effects found in the study. The second factor--control for test frequency--was negatively related to size of effect. Programs that did not control for test frequency (i.e., programs in which the control group took few or no unit tests) produced the strongest effect sizes. The third factor, natural science content, was also negatively related to effect size. Studies carried out in natural science subjects reported weaker effects than did studies in the social sciences and mathematics.

A fourth variable, type of remediation strategy used in a study, also contributed to the prediction of size of effect, but the regression coefficient for this variable reached only a borderline level of significance, $p = .052$. The tendency, however, was for studies using Keller's strategy to produce larger effects than were produced by in studies using Bloom's strategy.

Examination Performance and Student Aptitude

Six studies provided data on final examination performance for students at different ability levels. In all but one of these studies, the effect size was higher for less able students and lower for more able ones. The average effect size for the less able students in all six studies (M = .94) was also higher than it was for the more able students (M = 0.42).

Seven studies also reported correlations between achievement and aptitude for both mastery and nonmastery groups. In each of the seven studies, the correlation was higher in the nonmastery group, but only one of the 7 differences can be considered statistically significant. The average correlation between aptitude and achievement in all seven studies was .38 for the mastery group and .61 for the nonmastery group.

Variation in Examination Performance

Testing for mastery also reduced variation in student examination performance. Twenty of the 34 studies reported amount of variation in final examination scores for experimental and control groups separately, and in the majority of studies the variation in final examination scores was less for the experimental group. In the average study, variation of examination scores of the experimental group was 87% of the variation in scores of the control group.

Instructional Time

Although the experimental group students outperformed the students in the control group in most cases, the experimental group also required more instructional time. Eight studies reported on instructional time for experimental and control groups separately, and all eight showed increased instructional time for the experimental group. On the average, the experimental group required 25% more instructional time than did the control group.

Discussion

This meta-analysis provides support for some of the basic claims of mastery theory. It shows that testing for mastery has positive effects on student achievement. Mastery testing raised final examination scores in the typical study by 0.52 standard deviations, or from the 50th to the 70th percentile. Testing for mastery seemed to be equally effective for precollege and college students, and it seemed to work equally well in conventional classes and in individualized ones.

This study also demonstrated that effects of mastery testing are especially strong on low aptitude students. Because low aptitude students are affected more by mastery testing than are high aptitude students, a mastery requirement also has the effect of diminishing individual differences in student achievement in classes and reducing the correlation between aptitude and achievement. All these effects of mastery testing are predictable from the mastery model.

Effects of mastery testing were moderate to large in the typical study, and under certain conditions its effects were clearly large. For effects to be optimal, the study had to meet several criteria. First, the mastery level required in the study had to be high (95% to 100%). Second, the mastery demonstration had to be a formal one on actual tests. Assuming that students achieve mastery after meeting with a tutor, as Bloom does, reduces the potency of mastery procedures. Third, the study had to take place in mathematics or the social sciences. Investigations in the natural sciences yield less impressive results than do studies in these content areas. And fourth, the study had to lack a control for the number of tests taken by nonmastery and mastery groups. When nonmastery and mastery students are given the same number of tests, effects of mastery testing are smaller than when nonmastery students are given few tests and mastery students are given many.

In a number of important respects, however, the evidence in this review falls short of providing complete support for the claims of mastery theorists. First, effects of mastery testing were not so large as those predicted by mastery theorists. The two-sigma effect described by Bloom (1984) is far larger than the effect of mastery testing found in this meta-analysis: one-half sigma.

Nor does the evidence support Bloom's specific prediction that variation in performance will be reduced to near zero with mastery testing procedures. Most of the studies included in this meta-analysis reported reduced variation in student achievement with mastery testing, but no study reported zero or near-zero variation.

Finally, mastery testing increased substantially the amount of instructional time required in an area. And contrary to Bloom's predictions, instructional time differences for mastery and nonmastery classes were as great in longer studies as they were in shorter studies. As Arlin and Webster (1983) have suggested, mastery theorists may have to revise the time aspects of their theory to account for such results.

References

Arlin, M., & Webster, J. (1983). Time costs of mastery learning. Journal of Educational Psychology, 75, 187-195.

Block, J. H., & Burns, R. B. (1976). Mastery learning. In L.S. Shulman (Ed.), Review of Research in Education (Vol. 4). Itasca, IL: F.E. Peacock.

Bloom, B. S. (May, 1968). Mastery learning. Evaluation Comment (Vol. 1, No. 2). Los Angeles: University of California at Los Angeles, Center for the Study of Evaluation of Instructional Programs.

Bloom, B. S. (1984). The two-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13(6), 4-16.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage Publications.

Keller, F. S. (1968). "Good-bye teacher...." Journal of Applied Behavioral Analysis. 1, 79-89.

Table 1

Categories Used to Describe Studies

Treatment Features

- Remediation strategy used: Keller model and Bloom model
- Mastery level required: 70 - 100%
- Number of attempts allowed on unit tests
- Number of different forms of unit tests
- Type of unit test
- Duration of treatment

Methodological Features

- Random vs. nonrandom assignment
- Control for instructor effects
- Control for historical effects
- Control for test-author bias
- Control for bias in test scoring
- Control for test frequency
- Statistical control in outcome measurement

Setting Features

- Class level
- Course content
- Teaching method

Publication Features

- Document type
- Year of publication

Table 2

Major Features and Achievement Effect Sizes in 34 Studies on Mastery

Study	Place	Class level	Course content	Remediation strategy	Duration in weeks	Mastery level	Control for test frequency	Effect size
Arlin & Webster, 1983	Vancouver, Canada	7th	Sailing	K	1	90	Yes	0.53
Block, 1972	Portland, Oregon	8th	Math	K	1	95	Yes	0.64
Bostow & O'Conner, 1973	University of Florida	College	Psychology	K	10	90	Yes	0.82
Burrows & Okey, 1975	Bloomington, Indiana	4 & 5	Math	B	2	100	No	1.16
Caldwell, Bissonnette, Klishis, Ripley, Farudi, Hochstetter, & Radiker, 1978	West Virginia University	College	Psychology	K	16	93	Yes	1.07
Dunkelberger & Heikkinen, 1984	Greenville, Delaware	9th	Science	K	15	80	Yes	0.15
Dustin & Johnson, 1974	State University of New York, Plattsburgh	College	Psychology	K	8	100	Yes	0.76
Fagan, 1976	Dallas, Texas	7th	Social Studies	B	4	80	Yes	-0.11
Fehlen, 1976	University of Minnesota & Mankato State College	College	Math	K	10	90	Yes	0.34
Fernald, Chiseri, Lawson, Scroggs, & Riddell, 1975	University of New Hampshire	College	Psychology	K	7	90	Yes	0.20
Fiel & Okey, 1975	Bloomington, Indiana	8th	Math	B	2	100	Yes	0.47
Goldwater & Acker, 1975	University of Victoria, Canada	College	Psychology	K	15	80	No	0.65
Honeycutt, 1974	Ohio State University	College	Psychology	K	12	80	No	0.99
Hutchcraft & Kruger, 1977	Wright State University	College	Statistics	K	11	90	Yes	0.21
Hymel & Gaines, 1977	----------	7-9	Math	K	4	80	No	1.46
Hymel & Mathews, 1980	New Orleans, Louisana	7-9	Social Studies	B	2	100	Yes	0.99

Table 2 (continued)

Study	Place	Class level	Course content	Remediation strategy	Duration in weeks	Mastery level	Control for test frequency	Effect size
Jones, 1975	Savannah, Georgia	7th	Social Studies	K	4	85	Yes	0.18
Knight, Williams & Jardon, 1975	University of Oklahoma	College	Psychology	K	12	100	Yes	1.58
Komaridis, 1971	University of Nebraska	College	Psychology	K	2	90	Yes	0.31
Kulik, Kulik, & Hertzler, 1977	University of Michigan, Dearborn	College	Biology	K	12	75	Yes	0.38
Lawler, 1971	Florida State University	College	Education	K	10	80	Yes	0.31
Leppmann & Herrmann, 1981	University of Guelph, Canada	College	Psychology	K	32	80	Yes	0.42
Lewis, 1984	Memphis State University	College	Math	K	12	80	No	0.57
Martin & Srikameswaran, 1974	University of West Ontario, Canada	College	Chemistry	K	32	75	No	0.28
Meznarich, 1978	Texas A&M University	College	Engineering	K	13	70	Yes	-0.42
Myers, 1976	West Georgia College	College	Geography	B	11	85	Yes	0.36
Nation, Knight, Lamberth, & Dyck, 1974	University of Oklahoma	College	Psychology	K	8	100	Yes	0.90
Nation, Massad & Wilkerson, 1977	Texas A&M University	College	Psychology	K	8	90	Yes	0.39
Nation & Roop, 1975	Texas A&M University	College	Psychology	K	16	100	Yes	0.32
Okey, Brown, & Fiel, 1972	University of California, Santa Barbara	College	Science	B	--	100	No	0.30
Semb, 1974	University of Kansas	College	Psychology	K	6	100	Yes	0.80
Sheldon & Miller, 1973	Southern California community colleges	College	Math	B	17	75	No	0.28
Slavin & Karweit, 1984	Philadelphia, Pennsylvania	9th	Math	B	26	80	Yes	0.01

Table 2 (continued)

Study	Place	Class level	Course content	Remediation strategy	Duration in weeks	Mastery level	Control for test frequency	Effect size
Wentling, 1973	Illinois	Senior High	Science	K	5	80	Yes	0.38

Table 3

Means and Standard Errors of Effect Sizes for 32 Mastery Learning Studies Classified by Study Feature

Study feature	N	Effect size	
		M	SE
Remediation strategy			
Keller	26	0.54	0.08
Bloom	8	0.46	0.15
Unit mastery level***			
70-80	14	0.40	0.12
81-90	9	0.35	0.05
91-100	11	0.82	0.11
Duration of treatment			
Up to 4 weeks	9	0.63	0.17
5 to 8 weeks	6	0.57	0.12
9 to 12 weeks	9	0.60	0.15
13 to 16 weeks	5	0.35	0.25
17 weeks or more	4	0.31	0.11
Attempts allowed on unit tests			
One or two	12	0.54	0.12
Three	10	0.52	0.13
Four or more	12	0.51	0.15

Table 3 (continued)

Study feature	N	Effect size M	Effect size SE
Test forms used for repeating unit tests			
Same form	8	0.53	0.18
Different form	24	0.47	0.08
Kind of item on unit tests			
Objective	29	0.51	0.08
Essay	4	0.69	0.17
Subject assignment			
Random	15	0.61	0.09
Nonrandom	18	0.46	0.11
Control for instructor effect			
Same	22	0.47	0.07
Different	12	0.61	0.16
Evaluation design			
Post test only	16	0.64	0.09
Pre-post or covariance adjusted	18	0.42	0.11
Control for test frequency*			
Yes	26	0.45	0.08
No	8	0.74	0.15
Control for evaluator involvement			
Involved	21	0.47	0.07
Not involved	13	0.60	0.16

Table 3 (continued)

Study feature	N	Effect size M	Effect size SE
Class level			
Pre-college	14	0.51	0.12
College	20	0.53	0.09
Course content*			
Math	9	0.60	0.15
Science	6	0.18	0.12
Social science	18	0.60	0.10
Teaching method			
Conventional	16	0.54	0.11
Individualized	18	0.51	0.10
Source of study**			
Unpublished	7	0.79	0.17
Dissertation	7	0.17	0.12
Published	20	0.55	0.08
Year of report			
Up to 1974	13	0.54	0.07
1975-1979	15	0.54	0.14
1980 and after	6	0.44	0.14

*$p < .10$.

**$p < .05$.

***$p < .01$.

Table 4

Multiple Regression Analysis of Achievement Effect Sizes

Independent variable	Partial correlation	Regression coefficient	Probability
Unit mastery (Range: 70 to 100)	.383	0.015	.036
Remediation strategy (1 = Bloom, 2 = Keller)	.359	0.276	.052
Test frequency for control group (1 = no unit test, 2 = midterms only 3 = same as experimental group)	-.523	-0.279	.003
Dissertation (1 = no, 2 = yes)	-.296	-0.242	.112
Natural science content (1 = no, 2 = yes)	-.405	-0.256	.027